Received: 6 May 2023 | Accepted: 29 September 2023

DOI: 10.1111/twec.13522

ORIGINAL ARTICLE

The World Economy WILEY

Labour-saving automation: A direct measure of occupational exposure

Fabio Montobbio[1,2,3] | **Jacopo Staccioli**[1,4] | **Maria Enrica Virgillito**[1,4] | **Marco Vivarelli**[1,5,6]

[1]Department of Economic Policy, Catholic University of the Sacred Heart, Milano, Italy

[2]BRICK, Collegio Carlo Alberto, Torino, Italy

[3]ICRIOS, Bocconi University, Milano, Italy

[4]Institute of Economics, Sant'Anna School of Advanced Studies, Pisa, Italy

[5]Maastricht Economic and Social Research Institute on Innovation and Technology, UNU-MERIT, Maastricht, The Netherlands

[6]Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA), Bonn, Germany

Correspondence
Maria Enrica Virgillito, Institute of Economics, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, 56127 Pisa, Italy.
Email: mariaenrica.virgillito@santannapisa.it

Funding information
Ministero dell'Università e della Ricerca

Abstract

This article represents one of the first attempts at building a *direct* measure of occupational exposure to robotic labour-saving technologies. After identifying robotic and labour-saving robotic patents, the underlying 4-digit CPC (Cooperative Patent Classification) code definitions, together with O*NET (Occupational Information Network) task descriptions, are employed to detect functions and operations which are more directed to substituting the labour input and their exposure to labour-saving automation. This measure allows us to obtain fine-grained information on tasks and occupations according to their text similarity ranking. Occupational exposure by wage and employment dynamics in the United States is then studied, and complemented by investigating industry and geographical penetration rates.

KEYWORDS

labour markets, labour-saving technology, natural language processes, technological unemployment

1 | INTRODUCTION

Robots are coming! Statements such as this have become a mantra in recent years, together with the perception that "This time is really different" (Brynjolfsson & McAfee, 2012, 2014; Ford, 2015).

Much literature on the effects of the new wave of automation on human labour has been produced since then. Indeed, the pervasiveness of such new technological artefacts has been one of the most relevant aspects portending troublesome scenarios; among the most radical authors, Frey and Osborne (2017) suggest that 47% of total US employment is associated with occupations that are potentially automatable, a very much debated figure which has been revised downwards by further estimations (Arntz et al., 2016) giving a figure of 9% when looking at tasks rather than occupations. Nedelkoska and Quintini (2018), performing an analysis across 32 OECD countries, also reveal a large degree of cross-country variability, with estimated automation probabilities for the median job ranging between 39% and 62%. Recent empirical evidence tends however to agree that low- and medium-skilled workers, mainly, executing routinised tasks, are particularly at risk (Acemoglu & Restrepo, 2018, 2019, 2020a; Autor & Dorn, 2013; Frey & Osborne, 2017). At the same time, while some papers find a negative impact on employment and wages, systematic evidence of the labour market impact of robotic technologies remains elusive (Calvino & Virgillito, 2018; Mondolo, 2022).

In the literature, the unfolding of the impact of robotics on labour markets, in terms of occupations and wages, has mainly been estimated by two alternative methods. The first method is based on experts' judgement on a subset of occupations, expanded over the entire occupational structure by a classifier-system algorithm (e.g., Frey & Osborne, 2017; Nedelkoska & Quintini, 2018). The second approach has been leveraging robotic adoption at the sectoral level, relying on the International Federation of Robotics dataset and looking at the impact on local labour markets (e.g., Acemoglu & Restrepo, 2018, 2019, 2020a).

Currently, a direct measure of human substitutability and occupational exposure, ideally based on the functions and operations executed by labour-saving (LS) technologies, is still absent (see Section 2). We contribute to this literature and provide a direct link between human tasks and machine functions and, as a result, quantify occupational exposures to LS innovation in robotics. In doing so, we build a new measure of similarity between the textual description of the tasks performed in an occupation and the functions performed by observed robotic LS innovations.

First, we leverage the identification of robotic LS technologies by means of natural language processing on robotic patents (Montobbio et al., 2022) and we then perform a task-based textual match between the descriptions of technological classifications (the so-called CPC codes) attributed to robotic LS patents and the O*NET dictionary of occupations. The match exploits a cosine-similarity matrix that measures the proximity of the two dictionaries of words. The first result of our study is therefore the construction of a direct measure of similarity between a dictionary of technological LS functions and a dictionary of human-based functions. This is a methodological advancement to measure proximity between humans and machines and allows us to derive a direct measure of exposure.

In the second step, we aggregate tasks into occupations and derive a measure of exposure of each task and related occupations to robotic LS technologies. We find that the distribution of the similarity scores across tasks and occupations is very skewed, with high-similarity events being quite rare, given the underlying heterogeneity between the two text corpora. Nonetheless, restricting the analysis to the top decile of the similarity distribution, around 8.6% of the overall US employed workforce (approximately 12.6 million jobs) is at risk of substitutability. The most affected occupations are “Material Moving Workers”, “Vehicle and Mobile Equipment Mechanics, Installers, and Repairers”, “Other Production Occupations”. Logistics and production activities are those most exposed to LS technologies, in line with the evidence that among the top owners of LS patents, Amazon and UPS stand out (Montobbio et al., 2022). However, among the top patent holders, a quite diverse range of firms is visible, e.g., Boeing, Seiko, Canon, Hyundai Motors. The logistics segment is therefore targeted not because it reflects the sector of belonging of patent owners, but rather because it is considered easier to automate by means of robots within different industries. In line with the lean production paradigm, logistics activities have to be minimised and eventually completely automated.

To validate our methodology, we perform a robustness analysis by replicating the text similarity exercise between robotic LS patents' full texts and the same O*NET task descriptions (see Appendix 3). The patent-task match, when full texts are used but LS functions are not (through CPC code descriptions, see above), correctly pinpoints those occupations developing new innovative robotic technologies and their systems of adoption (e.g., Robotics Engineers; Robotics Technicians). This result reinforces the goodness of our procedure because it shows that it distinguishes substitutability detected via more prevalent functions in LS patents from complementarity detected via the match with the entire patent text.

Then, we link the similarity measure to the actual US labour market in terms of occupations and wages. We match our data to the Occupational Employment and Wage Statistics (OEWSs) from the US Bureau of Labor Statistics for 8-digit SOC occupations (1999–2019). Regression estimates present a monotonically negative relationship between occupational exposure and both (i) wage level and growth and (ii) employment growth. Remarkably, the expected U-shaped pattern (Acemoglu & Autor, 2011) is found neither in wages nor in occupational growth. In other words, cutting-edge robotic innovative efforts look to be directed towards the weakest and cheapest, and not the middle, segment of the labour market. Finally, a geographical breakdown across US states shows that the Rust Belt area, the region surrounding the Great Lakes experiencing industrial decline, and the South-East area, with a higher prevalence of African-American communities, record the largest employment shares of occupations that are particularly exposed to robotic LS technologies.

Our results highlight the fact that the drivers of labour-saving innovations are not those theoretically expected according to a theory of labour substitution induced by higher prices of labour inputs. Indeed, labour-saving innovative efforts are more directed at substituting and automating the cheapest segment of the labour market. Nonetheless, some of the highly targeted occupations, like Transportation and Material Moving, are still growing in terms of employment share. Therefore, the overall labour market dynamics cannot be defined uniquely by the direction of labour-saving technical change but rather by the interaction between demand-driven structural change and supply-driven technical change. Again, the archetypical case is Amazon, which represents the second US employer and whose massive share of the labour force involves Transportation and Material Moving operators. The latter occupations are targeted by LS efforts but still represent the fourth occupation by employment share in 2019.

The remainder of the article is organised as follows: Section 2 discusses the literature and evidence available; Section 3 presents the datasets used and Section 4 the adopted methodology; Section 5 shows and discusses our results, presenting task and occupational exposure, labour market, industry and geographical penetration rates. Section 6 concludes the article.

2 | STATE OF THE ART

As briefly mentioned in the introduction, the effect of robotic applications on labour markets, in terms of occupations and wages, has been estimated in previous research mainly by using two alternative methods. The first method is that introduced by Frey and Osborne (2017), who constructed an automation probability starting from experts' judgements on a subset of 70 occupations, and then expanding the evaluation over the entire occupational structure by means of a classifier-system algorithm. Experts were asked about the probability of automating some particular human functions. This approach was then employed by Nedelkoska and Quintini (2018) to study 32 OECD countries using the PIAAC dataset and was revised downwards by Arntz et al. (2016).

The second approach involves leveraging robotic adoption at the sectoral level, relying on the International Federation of Robotics dataset and looking at impacts on local labour markets. This is the route taken by Acemoglu and Restrepo (2018, 2019, 2020a) who generally predict that a higher number of robots per employee decreases wages and occupations for low-wage workers. However, cross-country studies at the industry level find a positive impact of robotic adoption on labour productivity and less clear-cut evidence on employment reduction. For instance, while Chiacchio et al. (2018) find results very consistent with Acemoglu and Restrepo (2020a), Dauth et al. (2021) conclude that robots do not significantly reduce total employment, although they do reduce the low-skilled workers' employment share, particularly in manufacturing.

Shifting to studies using firm-level data, the results are conflicting. Domini et al. (2021), using robotic adoption or, alternatively, imported capital equipment, do not detect labour expulsion but rather employment growth. Interestingly enough, in some studies the positive employment impact at the firm level appears entirely due to the so-called "business stealing effect"—i.e., innovative adopters gain market share at the expense of non-innovators (Dosi & Mohnen, 2019)—since negative employment impacts emerge once non-adopters and sectoral aggregates are taken into account (see Acemoglu et al., 2020; Koch et al., 2021).

More recent papers have focused on artificial intelligence, the purportedly newcomer disruptive technology, often blamed for having a strong LS impact on white-collar jobs, more related to service activities. Felten et al. (2021), who refine the measure proposed in Felten et al. (2018), link the Electronic Frontier Foundation dataset (EFF), within the AI Progress Measurement initiative, with O*NET (abilities). A direct matching between 10 AI-selected scopes of application (abstract strategy games, real-time video games, image recognition, visual question answering, image generation, reading comprehension, language modelling, translation and speech recognition) and human abilities is conducted. The matching is performed by crowdsourcing a questionnaire to gig workers at Amazon's Mechanical Turk (mTurk) web service. The questions administered to 2000 mTurkers residing in the United States asked whether they believed that AI application is related to, or could be used for, each of the 52 abilities listed in the O*NET. The study reports higher AI exposure for white-collar workers. However, the measure gives no information on any direct replacement or complementarity effect.

Webb (2020) proposes a direct measure of exposure via co-occurrence of verb-noun pairs in the title of AI patents and O*NET tasks. However, titles of patents are hardly informative regarding the underlying functions executed by the technological artefact, and restricting the study to verb-noun pairs has a high likelihood of false positives. The measure of exposure is not constructed in terms of the overall similarity of the two corpora but rather in terms of the relative frequency of occurrence of the elicited pairs in AI titles versus the remaining titles of non-AI patents. Moreover, the proposed methodology does not permit distinguishing labour-saving from labour-augmenting technologies.

Acemoglu et al. (2022) look at AI-exposed establishments and their job posts using Burning Glass Technologies data, which provide wide coverage of firm-level online job postings, linked to SOC occupational codes. To account for the degree of firm-level AI exposure, three alternative measures are employed, namely, those proposed by Brynjolfsson et al. (2018), Felten et al. (2021) and Webb (2020). Not surprisingly, considering adoption at a still relatively niche level, no clear effects at the industry and occupational levels are detected, while recomposition towards AI-intensive jobs is spotted. In addition, they do not find evidence of any direct complementarity between AI job posts and non-AI jobs, hinting therefore at a prevalent substitution effect and workforce recomposition, rather than a productivity-enhancing effect of AI adoption.

The closest analysis to our own is that performed by Kogan et al. (2021), who construct a text-similarity measure between the corpus of so-called *breakthrough innovations*, according to the methodology devised by Kelly et al. (2021), and the fourth edition of the Dictionary of Occupation Titles (DOTs). The measure is constructed to allow for time variability by keeping constant the textual content similarity but summing it for each defined breakthrough innovation at each time step, exploiting patent information over the period 1850–2010. Breakthrough innovations, identified as the distance between backward and forward similarity of each filed patent compared to the existing stock of patents, are by no means ex-ante defined as being labour saving in nature. In addition, the way the measure is built reflects more the dynamics of breakthrough innovations according to their emergence along subsequent technological revolutions, quite akin to the findings of Staccioli and Virgillito (2021), rather than the actual penetration of these technologies in the labour market. Therefore, what the measure captures is more the clustering of technologies under mechanisation in the first period of analysis, followed by automation and the ICT phase. They find that most exposed occupations lost in terms of wages and employment level and that over time white-collar workers became relatively more exposed compared to blue-collar ones. However, it is not clear whether the results are reflecting more long-run dynamics in technological and structural change rather than actual similarity between patents and occupations. Indeed, the within patent-occupation text-similarity is kept constant over time.

The measure defined by Kogan et al. (2021) has been applied by Autor et al. (2022), who were interested in devising the entry of new work titles over time in the historical records of the so-called Census Alphabetical Index of Occupations (CAIs), an index listing all new work-title entries. The authors define complementary technologies those patents matched with the CAI text (new job titles), and labour-saving technologies as the ones linked to the DOT text (existing job titles). The article documents the increasing entry of white-collar middle-paid occupations in the period 1940–1980, while since 1980 new jobs have been concentrated in services provided by both the highly educated and the less-educated. Another application of the Kogan et al. (2021) measure was with reference to I4.0 patents in Meindl et al. (2021), the authors matching in this case the patent text corpus with the "detailed work activities" (DWAs) section of the O*NET. According to their results, financial and professional occupations are more exposed to I4.0 patents compared to non-I4.0 patents.

Table 1 presents a summary of the most relevant contributions discussed so far, with their methodologies and findings. With respect to the extant literature we advance along the following lines: first, we construct a direct similarity measure which is able to assign a specific value to the similarity across two dictionaries of words, respectively, covering the realm of technology (CPCs) and human functions (O*NET); second, rather than relying on patent titles and co-occurrences of specific verb–noun pairs (Webb, 2020), we extend a far more complete and accurate specification of technological content of patent titles to the entire dictionary of functions described in CPCs; third, by employing the CPCs classification rather than patent texts (Kogan et al., 2021; Meindl et al., 2021), we are able to create a matrix of similarity with every underlying technology, permitting the generalisation of our measure beyond specific robotic technology. In addition, we also avoid excluding the majority of textual content present in patent text which is clearly not

TABLE 1 Synoptic table of companion literature. Jobs refer to tasks aggregated at the occupational levels.

Contribution	Measure (automation and AI)	Level of analysis	Highest exposure
Frey and Osborne (2017)	Delphi method to identify 70 most-exposed occupations to automation, and machine-learning algorithm to cover the remaining ones	Occupations	Routine activities (Offices and Administrative Support; Sales and related, Service)
Arntz et al. (2016) and Nedelkoska and Quintini (2018)	Technological bottlenecks identified by Frey and Osborne (2017)	Tasks	Low-skill occupations
Acemoglu and Restrepo (2018, 2019, 2020a)	Share of robot adoption	Industry	Low-wage workers
Felten et al. (2018, 2021)	Questionnaire on 10 AI-selected scopes of application crowdsourced to mTurk workers	Jobs	White-collar workers
Webb (2020)	Co-occurrence of verb-noun pairs in the title of AI/robot/software patents and O*NET tasks	Jobs	Low-wage occupations to robots. Medium-wage occupations to software. High-wage occupations to AI
Kogan et al. (2021)	*Term frequency-inverse document frequency* matrix of the patent text of breakthrough innovations and DOT	Jobs	Time-varying exposure to occupations reflecting waves of technological change
This paper	*Term frequency-inverse document frequency* matrix of CPCs and O*NET tasks	Jobs	Low-wage occupations concentrated in production, installation and maintenance segments but also affecting service based activities (e.g., healthcare practitioners), geographically located in the ex-industrial areas and the South of the US

consistent with the description of human functions. Fourth, we make a methodological advance with respect to sheer co-occurrences, by using a more advanced technique of textual similarity, resilient to the numbers of words and weighting scarcer rather than abundant information; fifth, by assigning a weight to CPCs according to their labour-saving traits, we are able to distinguish neatly between labour-saving and labour-complementary technologies; sixth, we do not rely on subjective validation methods, such as outsourced questions to workers in Amazon mTurk (Felten et al., 2021) or alternatively to Delphi approaches (Frey & Osborne, 2017), but rather employing a state of the art technique in natural language processing.

3 | DATA DESCRIPTION

The first dataset used in this study is the O*NET, a primary source of occupational information, widely employed in the literature, on the actual content of workplace activities (Handel, 2016). It provides details for the US occupational structure at the 8-digit level. The O*NET content model allows researchers to gather information on a series of job attributes, namely, executed tasks, task ratings (in terms of importance, relevance, and frequency), abilities, education, training and experience, knowledge, skills, work activities, work context, work styles.

Among the many available descriptors, the detailed Task Statements descriptor contains specific definitions of the tasks performed by each 8-digit occupation, while the Task Statement Ratings permits the gathering of information on the actual importance, relevance and frequency of each task, each occupation being composed of a multiplicity of tasks, also variable across occupations. The definitions of *core* or *supplemental* tasks synthesise the numerical rating scores, as detailed below.

Let us take as an example the occupation 19-3011.00 defining "Economists". The latter perform 11 core tasks, some of them more specific to the occupation in itself (e.g., task 7537: "Develop economic guidelines and standards and prepare points of view used in forecasting trends and formulating economic policy"); other tasks are less occupational-specific (e.g., task 7542: "Supervise research projects and students' study projects"); yet others are considered as supplemental tasks (e.g., task 20051: "Provide litigation support, such as writing reports for expert testimony or testifying as an expert witness"). Such granular information provides the basis for constructing the dictionary of words defining human functions.

The second dataset employed is the Occupational Employment and Wage Statistics (OEWS) retrieved from the US Bureau of Labor Statistics. This dataset permits analysis of the evolution of employment dynamics, excluding the self-employed, and is directly linkable to the O*NET dataset via the SOC full-digit occupational codes. In addition, the dataset permits the extraction of information on the average and median nominal wages for each 8-digit SOC category.

Figure 1 is a snapshot of the US occupational structure in 2019, showing the ranking of occupational categories aggregated at 2-digit levels (22 codes, excluding military-specific occupations) in terms of employment shares, and their evolution over the last two decades. In 2019, the largest occupational share (14%) is populated by "Office and Administrative Support" workers; "Healthcare Practitioners", "Technical, Business and Financial Operations" and "Management" stand at 6%, while "Life, Physical, and Social Science" are less than 1%. The snapshot recounts the remarkable deindustrialisation of the US economy, with a prevalence of administrative operations but also of service activities related to the satisfaction of social needs, such as "Food Preparation and Serving Related", "Transportation and Material Moving", while "Production" is relegated to fifth position with less than 7% of the US workforce. The occupation showing the most growth is "Healthcare Support" with an almost 100% increase in employment share, followed by "Business and Financial Operators" and "Computer and Mathematical" occupations. In general, a negative relationship between employment share growth in the last 20 years and employment share levels in 2019 is detectable, with those occupations recording the highest shares also experiencing strong contraction, such as "Office and Administrative Support" or contraction, as for "Sales and Related". "Production" workers experienced the largest decline in employment share (−4%), a further confirmation of the accelerated deindustrialisation process in the US.

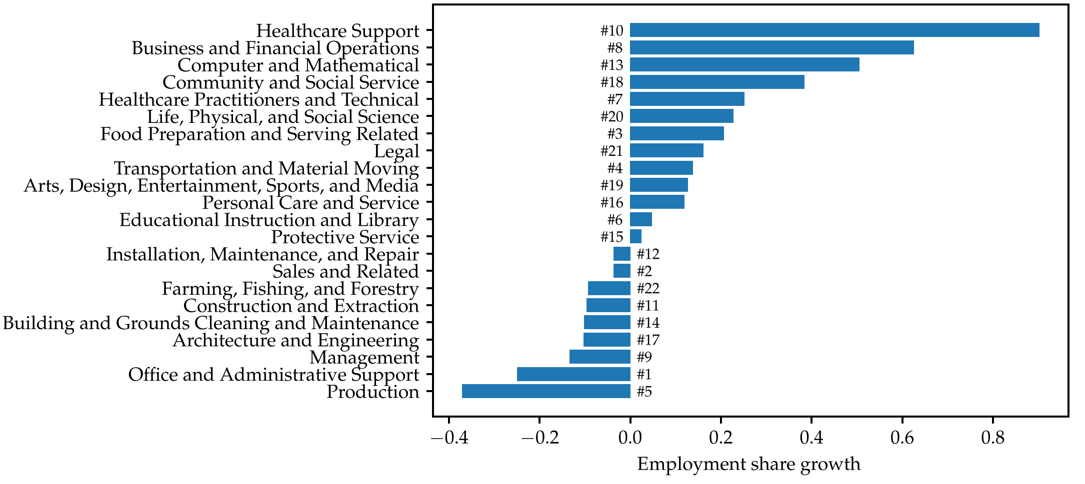

FIGURE 1 Employment shares growth between 1999 and 2019 for 2-digit SOC occupations (bars) and their relative ranking in 2019 (text).

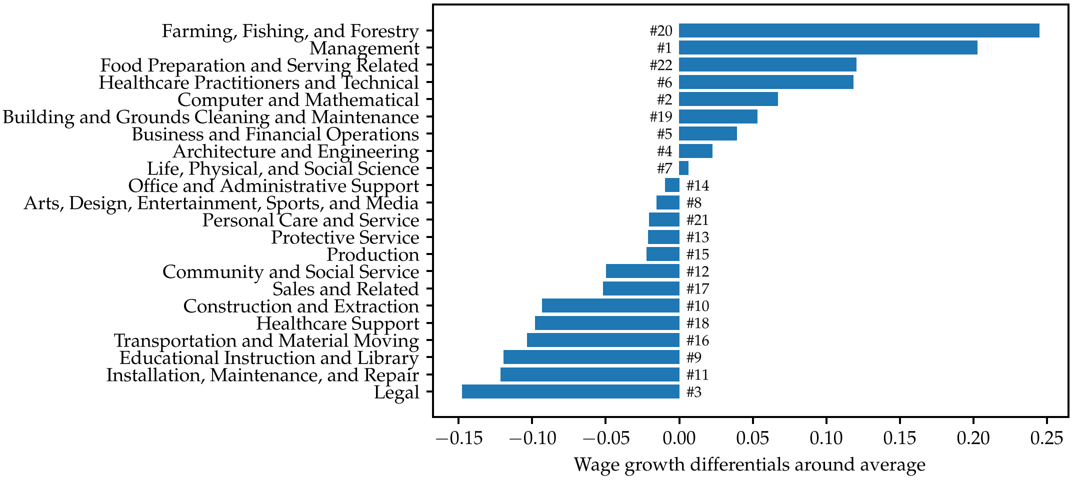

FIGURE 2 De-meaned wage growth between 1999 and 2019 for 2-digit SOC occupations (bars) and their relative ranking in 2019 (text).

Figure 2 presents the demeaned dynamics of median nominal wage growth by occupational categories at the 2-digit level, and the corresponding ranking of each occupation in terms of wage level in 2019. Not surprisingly, a hierarchical distribution of median wages emerges, with managerial median wages five times higher than food preparation activities. Albeit all occupations having experienced generalised nominal wage growth, with a minimum 50% increase, the occupations with the strongest growth in terms of remuneration have been both those which are top-paid and those which are bottom-paid according to the 2019 wage level, for example, the contrasting dynamics of managerial versus legal activities, in top remuneration tiers: for the former remuneration almost doubled when compared to 1999 levels, and increased by 20% more than average, while the latter experienced the lowest median wage increase with respect the average. Regarding the lower tiers, “Food Preparation and Serving Related” activities, the lowest-paid occupations in 2019, have likely experienced wage growth over 20 years and a wage increase 10% higher than the average, but still maintain their relative position in the wage hierarchy. Such a crystallised hierarchical structure reveals a more rigid than expected US labour market, wherein notwithstanding occupational changes in relative employment shares, the wage distribution across occupational categories has been quite sticky over the 20-year period.

The third dataset is that of USPTO patent applications in robotic technologies in the period 2009–2018, the most recent phase experiencing a steep increase in patenting activity in this field. As we shall describe below, both patent text corpora and technological fields are taken into account (CPC classification at 4-digit level).

4 | METHODOLOGY

In the present section, we explain the methodology necessary to build a new measure of similarity between the textual description of tasks performed in an occupation and the functions performed by LS innovations. In particular, we leverage the text similarity between the definitions of CPC (Cooperative Patent Classification) codes and the descriptions of tasks contained in the O*NET dictionary of occupations. Before delving into the methodological details of the text similarity measure that we devise (Section 4.2), it is useful to first summarise the relevant work by Montobbio et al. (2022) which brought about the set of LS patents constituting the starting point of the present analysis (Section 4.1). In Section 4.3, we explain how we map our measure of exposure from the task level to the occupation level. Our methodological workflow is summarised in the flowchart in Figure 3.

4.1 | Discovery of labour-saving patents

The contribution by Montobbio et al. (2022) in the study of robotic LS patents unfolds through three methodological steps. First, patents which either directly or indirectly relate to robotics technology are singled out. Second, a procedure is implemented in order to detect the underlying LS heuristics and pinpoint the set of explicitly LS patents. Finally, the most relevant CPCs in robotic LS patents vis-à-vis sheer robotic patents are identified. A brief technical summary of the relevant workflow is presented in Appendix 1.

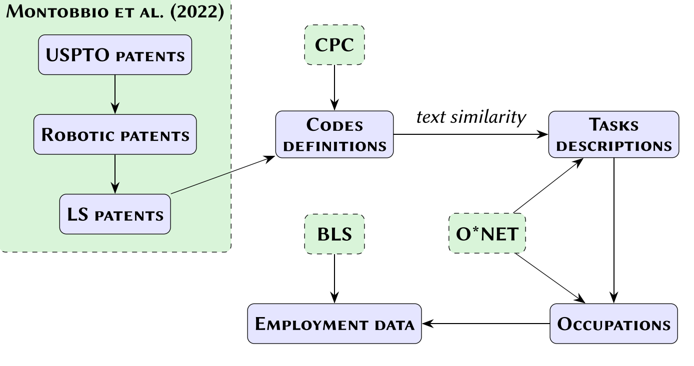

FIGURE 3 Flowchart of our methodology.

4.2 | Measuring exposure with text similarity

Within the scope of the present analysis, the technological content of LS patents is proxied by the official definitions of the relevant CPC codes. Patent publications are each assigned one or more classification terms indicating the subject to which the invention relates. The CPC system (we use the 2019.08 version) consists of 250,000+ distinct codes, organised according to a multi-level hierarchical structure. For the purpose of striking a fair balance between the density of information and granularity, we focus on 4-digit codes, of which 671 are present.

To match the technological content of LS patents to occupations, we rely on the O*NET database (we use version 25.1) which contains a thorough description of 19,231 distinct tasks, further aggregated into 923 8-digit SOC2018 occupations according to a weighting scheme detailed in Section 4.3. In the following, we aim to measure the pairwise text similarity of the 671 CPC code definitions and the 19,231 task descriptions.

From the methodological point of view (the formal description of which is given in Appendix 2), we adopt the so-called *bag-of-words* model and we measure textual proximity between CPC definitions and task descriptions by means of *cosine similarity* (see e.g., Aggarwal, 2018). The bag-of-words model entails the representation of text as a *multiset* of underlying words, which disregards any grammar structure and the order in which terms appear but keeps their multiplicity. The underlying assumption is that CPC-task pairs whose text consists of the very same words, possibly repeatedly, are more associated with one another than pairs which share few common words, or their frequency is negligible. As an example, consider the 4-digit CPC code A23F, whose definition reads (our emphasis).

> **COFFEE**; **TEA**; THEIR SUBSTITUTES; MANUFACTURE, **PREPARATION**, OR INFUSION THEREOF (**coffee** or **tea** pots A47G19/14; **tea** infusers A47G19/16; apparatus for making **beverages**, e.g. **coffee** or **tea**, A47J31/00; **coffee mills** A47J42/00),

and O*NET task 2209, whose definition reads

> **Prepare** and serve a variety of **beverages** such as **coffee**, **tea**, and soft drinks.

These two pieces of text have four words in common when their morphological roots are considered, namely ‘prepar*’, ‘beverage’, ‘coffee’ and ‘tea’, the latter two appearing multiple times in the CPC definition. As a consequence, they exhibit a high level of cosine similarity, ≈0.81 (note that the cosine similarity measure is bounded between 0 and 1). For the sake of comparison, consider the description of O*NET task 10209:

> Set up and operate machines, such as lathes, cutters, shears, borers, **millers**, grinders, presses, drills, or auxiliary machines, to make metallic and plastic workpieces.

The pair of CPC code A23F and O*NET task 10209 also exhibit a positive cosine similarity, although much lower (≈0.01), in that the two corpora only share the morphological root ‘mill*’. Had the two texts no words in common, the cosine similarity would be null (note that overly frequent English words such as ‘a’, ‘as’, ‘or’, ‘the’, etc., are discarded before computing the measure).

We compute the cosine similarity score for each pairwise combination of CPC codes and O*NET tasks to obtain a cosine similarity matrix S. The matrix S has dimension $671 \times 19{,}231$, one row for each CPC code and one column for each task, and each cell contains the similarity score of the underlying CPC-task pair. In total, there are 12,904,001 such pairs. The architecture of the matrix S is represented in Figure 4, where the task-occupation mapping, defined in the O*NET database itself, is also highlighted.

To make the cosine similarity matrix reflect the technological structure of the LS patents found by Montobbio et al. (2022), we multiply each row S of by the frequency of the associated CPC code in the whole set of LS patents. When a patent is assigned multiple CPC codes with the same 4-digit representation, each is taken into account separately. In other words, we filter the cosine similarity matrix by the distribution of technological codes across LS patents. In this way, we weigh the contribution of each CPC code to the occupational exposure of a certain task proportionately with how widespread the code is in LS patents. Finally, in order to rank O*NET tasks by similarity score with the ensemble of CPC codes of LS patents, we compute column sums of matrix across all CPC codes (rows). The result is a task similarity vector TS containing a unique measure of aggregate similarity to each task. Given a column vector C of frequencies of the 671 CPC codes among LS patents, the vector TS is defined simply as,

$$TS = S'C$$

where the usual matrix multiplication is intended. A ranking of tasks by (aggregate) similarity score is later presented in Table 2, where values have been rescaled between 0 and 1.

4.3 | From tasks to occupations

So far, we have measured the textual proximity of LS patents to each O*NET task, mediated by CPC codes, the result of which is stored in the task similarity vector (cf. Section 4.2). To draw conclusions about the effect of LS technologies on employment, we need to further aggregate the similarity measure at the occupation level.

The O*NET database defines 923 occupations as collections of underlying tasks, distinguishing between *core* and *supplemental* tasks. This classification takes into account three distinct measures, namely *importance*, *relevance* and *frequency*. To aggregate a task similarity measure to the relevant occupation in a sensible way, it is crucial to understand how the core/supplemental distinction is devised in the first place. Importance spans a range of between 1 and 5, while relevance and frequency are both represented as percentages. Core tasks are deemed critical to

Occupation	11-1011.00			···	53-7121.00		
Task / CPC	8823	8824	···	···	···	12809	12810
A01B	cos(A01B,8823)	cos(A01B,8824)	···	···	···	cos(A01B,12809)	cos(A01B,12810)
A01D	cos(A01D,8823)	cos(A01D,8824)	···	···	···	cos(A01D,12809)	cos(A01D,12810)
···	···	···	···	···	···	···	···
H05H	cos(H05H,8823)	cos(H05H,8824)	···	···	···	cos(H05H,12809)	cos(H05H,12810)
H05K	cos(H05K,8823)	cos(H05K,8824)	···	···	···	cos(H05K,12809)	cos(H05K,12810)

FIGURE 4 Architecture of the cosine similarity matrix S.

TABLE 2 Top 15 tasks by (rescaled) cosine similarity (rounded to the 2nd decimal digit).

Rank	Code	Description	Cosine similarity
1	14587	Load materials and products into machines and equipment, or onto conveyors, using hand tools and moving devices	1.0
2	3202	Move levers or controls that operate lifting devices, such as forklifts, lift beams with swivel-hooks, hoists or elevating platforms, to load, unload, transport, or stack material	0.96
3	3203	Position lifting devices under, over or around loaded pallets, skids or boxes and secure material or products for transport to designated areas	0.90
4	17928	Lift and move loads, using cranes, hoists, and rigging, to install or repair hydroelectric system equipment or infrastructure	0.89
5	15266	Manually or mechanically load or unload materials from pallets, skids, platforms, cars, lifting devices or other transport vehicles	0.88
6	14584	Remove materials and products from machines and equipment, and place them in boxes, trucks or conveyors, using hand tools and moving devices	0.86
7	11839	Transport machine parts, tools, equipment, and other materials between work areas and storage, using cranes, hoists or dollies	0.85
8	3217	Load materials and products into package processing equipment	0.85
9	12805	Operate conveyors and equipment to transfer grain or other materials from transportation vehicles	0.85
10	12323	Communicate with systems operators to regulate and coordinate line voltages transmission loads and frequencies	0.84
11	12798	Operate industrial trucks, tractors, loaders and other equipment to transport materials to and from transportation vehicles and loading docks, and to store and retrieve materials in warehouses	0.83
12	20387	Optimise photonic process parameters by making prototype or production devices	0.83
13	17496	Provide information about community health and social resources	0.83
14	13705	Unload materials, devices and machine parts, using hand tools	0.80
15	10757	Load, unload or adjust materials or products on conveyors by hand, by using lifts, hoists, and scoops or by opening gates, chutes or hoppers	0.80

the occupation; the criteria for a task to be classified as core require that relevance ≥67% and importance ≥3.0. Supplemental tasks are deemed less relevant and/or important to the occupation; two sets of tasks are included in this category, namely, tasks rated ≥67% on relevance but <3.0 on importance, and tasks rated <67% on relevance, regardless of importance.

Taking the O*NET definition of core and supplemental tasks into account, we impute task similarity to occupations with the following weights:

$$\text{core: } \frac{2/3}{\#\,\text{tasks in the occupation}}$$

$$\text{supplemental: } \frac{1/3}{\#\,\text{tasks in the occupation}}$$

It is worth noticing that occupations differ in the number of constituent tasks. This warrants rescaling the similarity contribution of each task accordingly, hence the denominator in the aforementioned weighting scheme.

5 | RESULTS

Our results are displayed below. This section is organised as follows: Section 5.1 presents the distribution of the text similarity measure across tasks and occupations. Section 5.2, exploiting the match with the OEWS dataset, permits an understanding of the degree of penetration of technological exposure in terms of labour market structure, looking at employment and wages, and industry composition. Finally, Section 5.3 presents the geographical distribution of LS threats.

5.1 | Task and occupational exposure

Table 2 shows the top fifteen tasks in terms of the similarity score. Notably, the tasks with a higher similarity score regard human functions related to the handling and moving of objects, materials and products, with the top three tasks being (i) “Load materials and products into machines and equipment, or onto conveyors, using hand tools and moving devices”; (ii) “Move levers or controls that operate lifting devices, such as forklifts, lift beams with swivel-hooks, hoists, or elevating platforms, to load, unload, transport, or stack material”; (iii) “Position lifting devices under, over, or around loaded pallets, skids, or boxes and secure material or products for transport to designated areas”. Such types of tasks are more prevalent in the logistics industry in that they require activities such as preparing boxes, packaging, sorting and routing items.

Figure 5 presents the distribution of the similarity measure across the entire set of 19,231 O*NET tasks, according to the Task Statements file. Given the wide spectrum of the tasks covered by the O*NET, it is not surprising that an extremely skewed distribution is obtained, with events of high similarity being extremely rare. Indeed, cosine similarity values beyond 0.8 apply to a tiny fraction of tasks and at this stage seem to constitute more the exception rather than the norm. However, such an extreme value distribution is quite comforting in terms of the reliability of the measure, the underlying information being quite sparse, inasmuch as the probability of false positives is low and the overall accuracy of the measure high. Nonetheless, one might consider the task domain irrelevant or less informative, given that occupations are defined not by a single,

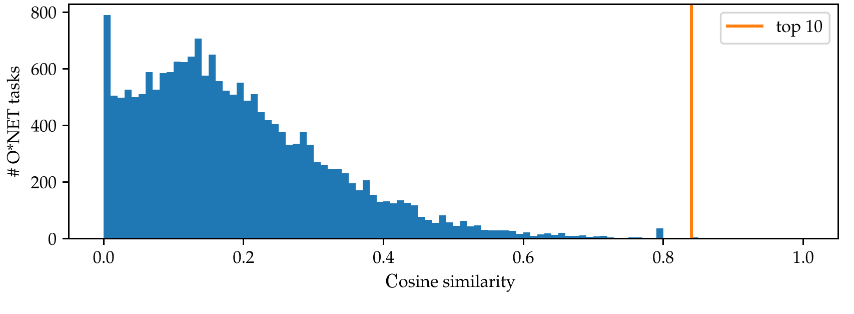

FIGURE 5 Distribution of the cosine similarity with respect to tasks.

but rather as a set of distinct tasks. The most exposed tasks therefore might turn out to be largely not core in many occupations, in line with the evidence provided in Figure 1, telling a picture of a labour market where administrative, office and sales occupations represent the largest shares of employed workers.

However, such a handful of tasks, when aggregated into occupations, happens to be quite revealing of the direction of robotic LS efforts. Table 3 shows the corresponding top 20 occupations most exposed to substitution by some form of automation or intelligent automation, obtained by aggregating tasks into occupations, as described in Section 4.3. Although the distribution by tasks is quite skewed, the very same information once aggregated at the SOC level, permits the detection of a series of occupations at strong exposure risk. The most exposed occupations are "Industrial Truck and Tractor Operators"; "Maintenance Workers, Machinery"; "Machine Feeders and Offbearers"; "Packers and Packagers, Hand". Such occupations are clearly among those reporting a higher incidence of tasks with higher similarity scores. From Table 3, a recurrent pattern emerges of some specific macro-occupational groups, as evidenced by the presence of the 2-digit occupations "Transportation and Material Moving" (53), "Installation, Maintenance, and Repair" (49), "Packaging and Filling Machine Operators and Tenders" (51), all recurrently ranking in the top twenty 8-digit most exposed occupations.

Table 4 pools the similarity measure, again rescaled between 0 and 1, for occupations at the 2-digit level. The occupational ranking aggregated at the 2-digit level presents the highest

TABLE 3 Top 20 occupations by (rescaled) cosine similarity (rounded to the 2nd decimal digit).

Rank	Code	Title	Cosine similarity
1	53-7051.00	Industrial Truck and Tractor Operators	1.0
2	49-9043.00	Maintenance Workers, Machinery	0.97
3	53-7063.00	Machine Feeders and Offbearers	0.94
4	53-7064.00	Packers and Packagers, Hand	0.91
5	49-2091.00	Avionics Technicians	0.87
6	51-9111.00	Packaging and Filling Machine Operators and Tenders	0.81
7	49-3041.00	Farm Equipment Mechanics and Service Technicians	0.81
8	49-3092.00	Recreational Vehicle Service Technicians	0.78
9	49-3042.00	Mobile Heavy Equipment Mechanics, Except Engines	0.77
10	47-2111.00	Electricians	0.76
11	49-9098.00	Helpers--Installation, Maintenance, and Repair Workers	0.75
12	49-9041.00	Industrial Machinery Mechanics	0.75
13	51-9082.00	Medical Appliance Technicians	0.75
14	47-3011.00	Helpers--Brickmasons, Blockmasons, Stonemasons, and Tile and Marble Setters	0.75
15	51-9191.00	Adhesive Bonding Machine Operators and Tenders	0.75
16	51-9023.00	Mixing and Blending Machine Setters, Operators, and Tenders	0.74
17	13-1032.00	Insurance Appraisers, Auto Damage	0.73
18	51-4111.00	Tool and Die Makers	0.73
19	49-9081.00	Wind Turbine Service Technicians	0.72
20	51-8013.04	Hydroelectric Plant Technicians	0.72

TABLE 4 2-digit occupations by (rescaled) cosine similarity (rounded to the 2nd decimal digit).

Rank	2-digit occupation	Cosine similarity
1	Production	1.0
2	Installation, Maintenance, and Repair	0.56
3	Construction and Extraction	0.54
4	Healthcare Practitioners and Technical	0.50
5	Architecture and Engineering	0.50
6	Transportation and Material Moving	0.43
7	Life, Physical, and Social Science	0.31
8	Education, Training, and Library	0.29
9	Office and Administrative Support	0.27
10	Management	0.26
11	Arts, Design, Entertainment, Sports, and Media	0.22
12	Computer and Mathematical	0.21
13	Business and Financial Operations	0.21
14	Personal Care and Service	0.13
15	Protective Service	0.12
16	Healthcare Support	0.11
17	Sales and Related	0.10
18	Farming, Fishing, and Forestry	0.07
19	Food Preparation and Serving Related	0.06
20	Community and Social Service	0.06
21	Building and Grounds Cleaning and Maintenance	0.03
22	Legal	0.0

similarity score in "Production" occupations which rank first, with a similarity score of 1. A substantial drop in the similarity score is visible in the second-ranking occupation (0.56). Notably, and differently from extant studies (Frey & Osborne, 2017; Nedelkoska & Quintini, 2018), exposure penetration not only targets the so-called standardised and routinised occupations. This result emerges with the presence in the top 10 most exposed occupations of "Healthcare Practitioners and Technical", "Architecture and Engineering", "Life, Physical, and Social Science", "Management". The latter activities are characterised by a low degree of routinised content and their high-ranking reflects that our measure of exposure also captures human functions entailing (i) movements in unstructured workplaces, such as those carried out by "Installation, Maintenance, and Repair Occupations" (ranking second) and by "Healthcare Practitioners and Technical" (ranking fourth), (ii) complex cognitive activities such as required in "Architecture and Engineering" (ranking fifth) and in "Life, Physical, and Social Science" occupations (ranking seventh) and (iii) social and relationship intelligence required in "Management" occupations (ranking tenth). Such a wide-ranging list of occupations provides a neat distinction when comparing our measure with routinisation indices, the share of robots at the industry level or even co-occurrences of robotic/AI patents with O*NET tasks, the latter able to target either low-wage occupations with reference to robots, or high-wage occupations with reference to AI (Webb, 2020). In addition, ranking the top ten occupations (although with a strong non-linear

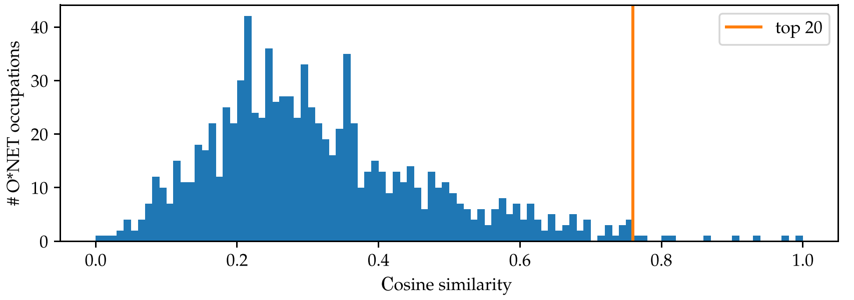

FIGURE 6 Distribution of cosine similarity with respect to 8-digit occupations.

threshold), covering from low- to high-wage jobs, reflects the weighting procedure assigned to tasks, which very much blurs the border between routinised versus non-routinised occupations, and assigns a positive similarity value additionally to those tasks which are not core but still are performed in a supplemental manner.

Albeit widespread, very high-level exposure is not common across disaggregated occupations. Figure 6 presents the histogram of occupations by similarity. Although less skewed than the task histogram, it confirms that high similarity is rare, affecting only a relatively small fraction of the entire occupational structure. The top 20 occupations with the highest similarity scores are indeed very few, and the mode of the distribution stands in the medium range of similarity, between 0.2 and 0.4. The area on the right of the orange bar identifies the range of the top-20 most exposed occupations.

However, if we move the bar to the left, Figure 7 (panel a) plots the quantile function of the similarity distribution in Figure 6. Given the highly skewed pattern, up to the eighth decile of the similarity distribution, there is a low range of variation, reaching the value of 0.4. Higher cosine similarity values, in the range [0.6, 1] occur only after the inflection point located around the ninth decile. This non-linear relationship reveals the existence of a threshold level beyond which exposure dramatically increases, while below this point exposure is reduced. Such threshold behaviour has a twofold implication: on the one hand, the high-exposure risk to substitutability affects a relatively tiny fraction of the entire occupational range, while on the other hand, whenever the risk is high, it accelerates rapidly, potentially leading to quite probable substitutability events.

Figure 7 (panel b), by using the O*NET–OEWS match, provides the effective number of employed workers per each occupation at risk of substitutability. As expected, the number of replaceable employees drops dramatically when the similarity value increases: the top decile of the similarity distribution, on the far-right, affects 8.6% of the employed working population, which amounts to approximately 12.6 million workers. Notably, and differently from other extant measures, our approach allows us to identify not a point value but rather an interval of exposure, which furthers understanding of how labour-substitutability hampers the labour force unevenly.

5.2 | Industry and labour market penetration

Occupations are distributed across industries, and therefore identifying those most and least affected is crucial for any potential policy intervention. Table 5 shows the relevance of occupational exposure to robotic LS technologies in each NAICS 2-digit sector by weighting the cosine similarity by the percentage of occupation membership in each sector. The measure, which takes a value 1 for the most part and a value 0 for the least exposed industry (in relative terms), depicts

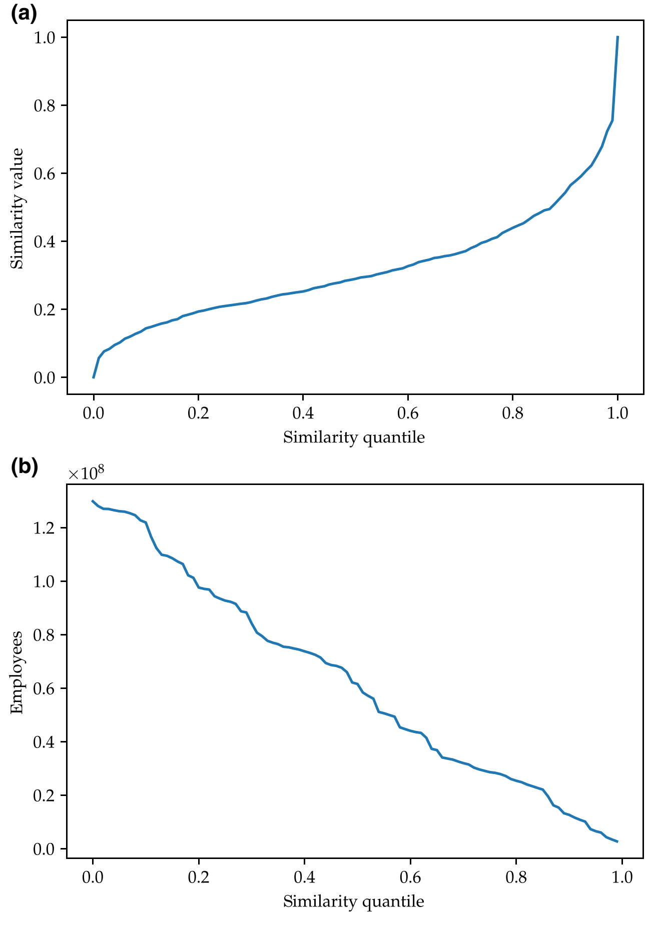

FIGURE 7 Quantile function of the similarity distribution for 8-digit occupations (top) and number of replaceable employees by quantile (bottom).

the manufacturing sector as the most exposed to automation. Not only does manufacturing head the ranking, but all the other sectors follow after a dramatic drop. Indeed, the industry ranking reflects the aggregation of scattered occupations entailing manual abilities and handling, as shown above, which however are largely concentrated in manufacturing. Moreover, manufacturing as an industry collects all those activities related to logistics and warehousing which are still currently under the parent manufacturing companies, while third-party logistics, carrying out logistics activities as a service, rank seventh across industries.

On the whole, robotic LS technologies will further the long-term ongoing deindustrialisation of the US economy (cf. Section 3). However, the high ranking of healthcare, social assistance and education, the most human-oriented industries, is remarkable. Such a result serves as a warning, already mentioned by Montobbio et al. (2022), regarding the direction of cutting-edge innovative efforts towards industries where, at least in principle, the human-based component should be preferable. Similarly, public administration ranks in the top five most exposed sectors to substitution, and this signals the ability of our measure to cover not only automation per se but also advanced digitalisation processes in administrative services. This is because labour-saving functions detected in our patents, and then mapped at the CPCs level, include not only automation technologies but also intelligent automation technologies. This is reflected by one-fourth of

TABLE 5 Relevance of exposed occupations to NAICS 2-digit sectors, obtained as a weighted average of similarity and occupation membership to the underlying sector, rescaled between 0 and 1.

Rank	2-digit sector	Cosine similarity
1	Manufacturing	1.0
2	Health care and social assistance	0.39
3	Education services	0.33
4	Construction	0.30
5	Public administration	0.21
6	Other services, except public administration	0.18
7	Transportation and warehousing	0.17
8	Retail trade	0.16
9	Professional, Scientific and Technical Services	0.11
10	Utilities	0.10
11	Administrative and support and waste management and remediation services	0.09
12	Information	0.09
13	Arts, entertainment and recreation	0.07
14	Accommodation and food services	0.07
15	Wholesale trade	0.07
16	Mining	0.06
17	Finance and insurance	0.05
18	Agriculture, forestry, fishing and hunting	0.04
19	Real estate and rental and leasing	0.02
20	Management of Companies and Enterprises	0.0

AI patents included in the LS patents. In addition, running the topic model over such patents, Montobbio et al. 2022 show that together with logistics activities, LS patents target human functions entailing interactive relations, patient treatments, and also activities such as learning and predicting. Such human functions are more prevalent in service-based occupations in healthcare, public administration and education. Notably, "Management of Companies and Enterprises", although recording a contraction in occupational employment shares in the last 20 years, presents the lowest similarity score.

The last battery of results is shown in Figures 8 and 9, presenting a non-parametric LOWESS estimate (locally weighted scatterplot smoothing) across O*NET occupations, as in Acemoglu and Autor (2011) and Webb (2020), of the relationship between the cosine similarity measure, employment, and wages, both in levels (2019) and growth rate (1999–2019). A neat negative, almost monotonically decreasing relationship emerges in three considered variables. Starting with employment levels (panel a), LOWESS estimates confirm the previous evidence, showing a low occurrence of high-similarity measures for the majority of employees. However, a negative relationship emerges when the similarity measure is compared against employment growth (panel b), revealing that shrinking occupations in the last two decades have also been those most exposed to robotic LS technologies. Such evidence confirms that, among other potential sources determining the displacement of some occupations, substitutionary technical change related to automation might have played a role. Alongside LOWESS, both OLS and LAD estimates are

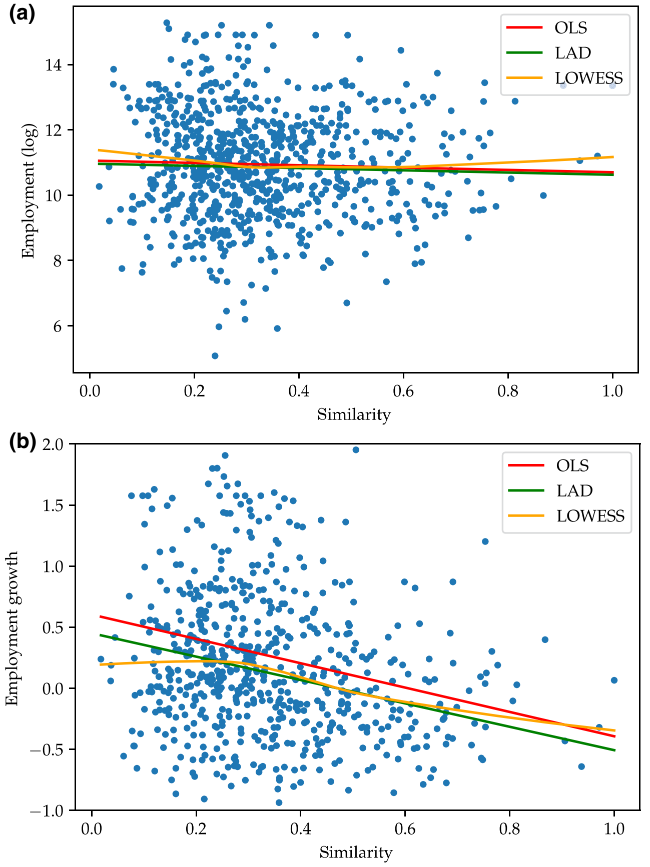

FIGURE 8 OLS, LAD and LOWESS estimates of the relationship between similarity and employment levels in 2019 (top) and 1999–2019 employment growth (bottom). A few extreme values are removed for visual convenience.

superimposed in Figures 8 and 9. Unlike OLS, LAD estimates the conditional median and is less sensitive to outliers in the data. Table 6 presents the coefficients of the similarity values for our four variables of interest. In line with the non-parametric estimation, the slopes are all negative and their signs are statistically significant, except for employment levels. The number of observations varies according to data availability in the OEWS dataset.

Which labour force segments are such innovative efforts directed towards? According to the LOWESS estimates, the relationship in terms of wage level and growth signals that the occupations most exposed to LS technologies are those which are lowest-paid and record the lowest wage growth. In other words, robotic LS technologies and their underlying patents are more directed towards substituting the cheapest segments of the labour force. This result is not consistent with the idea that a cheaper factor generates incentives to introduce technologies that use this factor more intensively. For example, Acemoglu and Restrepo (2018, 2019), in their task-based framework, suggest that automation process slows down when there is a lower effective cost of labour in the least complex tasks.

Anecdotal evidence suggests a high incidence of highly automatised production processes in already quite standardised workplaces (Ford, 2015); related to this, a large majority of case studies on Industry 4.0 questions the revolutionary content of the latest technological wave

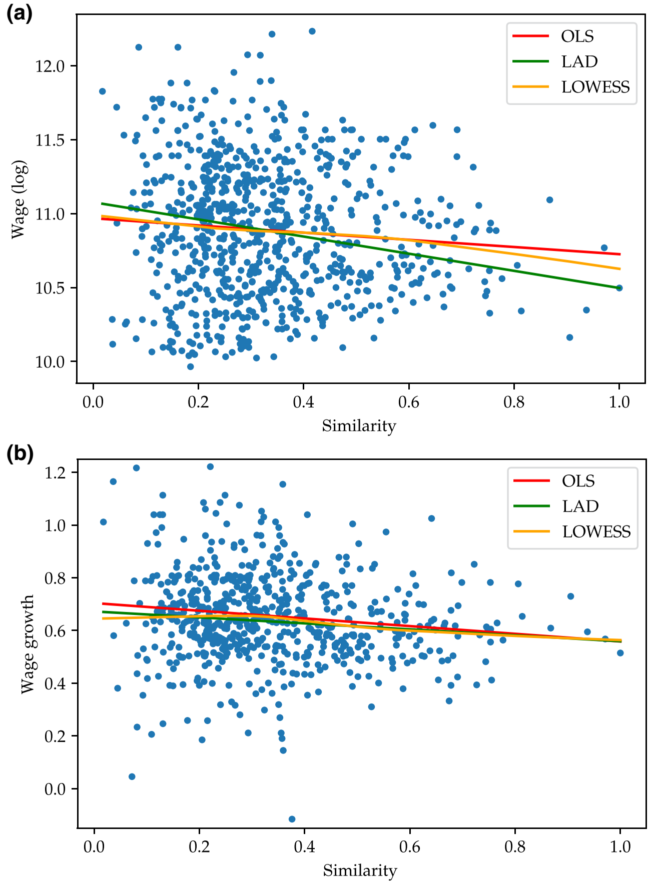

FIGURE 9 OLS, LAD and LOWESS estimates of the relationship between similarity and wage levels in 2019 (top) and 1999–2019 wage growth (bottom). A few extreme values are removed for visual convenience.

and highlights patterns of continuity with ICT (Cetrulo & Nuvolari, 2019; Cirillo et al., 2021; Krzywdzinski, 2021; Santarelli et al., 2022). High innovative efforts to automate cheap labour are what Acemoglu and Restrepo (2020b) define as "so-so" technologies. Far from judging labour from its remuneration, it is evident from our analysis that gains from automation in terms of productivity are sought in knowledge and technology, allowing for incremental upgrading of already automated processes and substituting the labour force therein involved. Indeed, our results point to a "furthering" of automation targeting production, manual and least-paid jobs, entailing both standardised but also non-standardised activities (such as moving objects in unstructured workplaces), therefore not fully consistent with the "routine-biased technical change" hypothesis (Acemoglu & Autor, 2011; Autor & Dorn, 2013) and at odds with a "Hicksian" direction of technical change induced by the level and change of costs of production factors.

5.3 | Geographical penetration

The United States are very much differentiated in terms of productive specialisation and ensuing occupational composition. Understanding the different geographical penetration of robotic LS

TABLE 6 Regression estimates.

Dependent variable	Method	Constant	Cosine similarity	# Observations
2019 employment (log)	OLS	11.0566***	−0.3577	804
		(0.133)	(0.362)	
	LAD	10.9629***	−0.3404	
		(0.179)	(0.489)	
1999–2019 employment growth	OLS	0.6034***	−0.9965***	659
		(0.065)	(0.173)	
	LAD	0.4506***	−0.9575***	
		(0.061)	(0.164)	
2019 wage (log)	OLS	10.9686***	−0.2431**	795
		(0.037)	(0.099)	
	LAD	11.0770***	−0.5802***	
		(0.052)	(0.141)	
1999–2019 wage growth	OLS	0.7041***	−0.1458***	655
		(0.018)	(0.048)	
	LAD	0.6723***	−0.1131**	
		(0.016)	(0.044)	

***p-value < .01; **p-value < .05 and *p-value < .1.

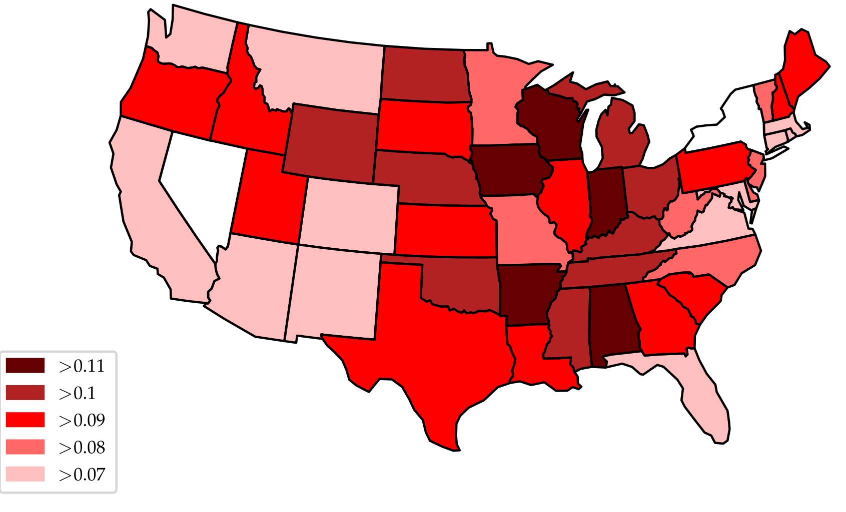

FIGURE 10 Disaggregation by state. Employment shares of most exposed occupations (top decile) to LS technologies. Continental US.

technologies across states is useful and informative, both as a validation exercise and as a tool to perform targeted policy actions. Figure 10 shows the state-level disaggregation of the top-ranking exposed occupations (top decile of the distribution in Figure 6). The heat map uses five colour shades, going from dark red for the states more exposed to LS technologies, to light red for less exposed states. The range of variation, considering the 8.6% share for the US as a whole, goes from 4% to 11%; in the figure states with a share below 7% are coloured white.

According to our results and going from the most to the least exposed states, the so-called Rust Belt area (i.e., the region surrounding the Great Lakes) is populated by darker reds, namely, Wisconsin, Indiana, Michigan, Illinois. Other dark red states are Alabama, Arkansas, Mississippi, Louisiana, located in the Southern East, where African-Americans are largely concentrated. Texas, with a prevalent fraction of Hispanic communities, has 9% of the occupational categories exposed to LS technologies. In contrast, all the states on the two coasts, the cradle of high-tech companies and of highly-specialised service activities are coloured in light red, i.e., Washington, California, Florida, Virginia. Notably, states such as New York and D.C. are white, the share of occupations exposed to LS technologies being below 7%.

From combined analyses of labour market trends, sectoral and geographical penetrations it emerges a high vulnerability of those segments of the labour market with a larger fraction of low-paid occupations in less developed states, with a higher concentration of ethnic minorities. In addition, although we do not consider age distribution due to the lack of data, other studies have pointed out that younger workers are more exposed to automation (Webb, 2020), although results are not univocal on that matter (Fillmore & Hall, 2021). Such findings suggest that occupational exposure to robotic LS technologies is negatively correlated with a proxy of economic performance at the geographical level. In this respect, looking at the joint dynamics of LS exposure and a proxy of economic performance might provide a reliable alert for policy makers in order to distinguish from potential positive effects of substitution in arduous and dangerous tasks. For example, the measure can be coupled with other information already present in the O*NET task descriptor, and correlate LS exposure with task exposure to health risks, not so differently from what has been done to monitor the possibility of teleworking when looking at information from performed tasks in outdoor spaces or entailing physical activities and manipulation of equipment (Dingel & Neiman, 2020). In this respect, potential multi-level analyses can be constructed, looking at the measure in conjunction with other metrics to help the policy maker target the most vulnerable occupations, or alternatively detect the extent to which substitution is carried out to improve safety at work.

6 | CONCLUSIONS

This article represents one of the first attempts at building a *direct* measure of occupational exposure to labour-saving technologies. We achieve this objective by making use of natural language processing (NLP) techniques. First, we leverage the information retrieved by Montobbio et al. (2022), allowing us to identify explicitly labour-saving robotic patent applications, collected in the 2008–2018 period. After identifying robotic and LS robotic patents, the underlying 4-digit CPC definitions are employed in order to detect functions and operations performed by technological artefacts which are more directed to substituting the labour input. Our measure allows us to obtain fine-grained information on tasks and occupations according to their similarity ranking. In addition, performing a match with occupational and wage statistics, we offer evidence on the relationship between our exposure measure, employment and wage levels, and their changes in the last 20 years. Industry and geographical penetrations are studied as well.

According to our results, the occupations which are more exposed to robotic labour-saving automation are those activities related to manual dexterity, manipulation, loading of objects into machines and equipment, lifting and load moving. These tasks mainly characterise occupations such as industrial truck and tractor operators, packaging and filling machine operators and tenders, and tool and die makers, but also include medical appliance technicians. When decomposing the information by industry, we find that manufacturing is the most exposed sector, followed by healthcare, social assistance and education services. In addition, although manufacturing is the most exposed sector, many identified tasks regard logistics activities.

LOWESS estimates of the relationship between occupational exposure and labour market variables confirm that robotic labour-saving innovations target low-paid occupations experiencing the lowest wage increases in the last 20 years. In addition, such low-paid occupations are also shrinking in terms of workforce. Although we cannot cover the entire set of potential confounding factors affecting such relationships, we find striking and clear-cut patterns of innovative efforts directed towards the cheapest and most vulnerable segments of the working population.

Exposure to LS automation and employment dynamics do not always go hand in hand. For instance, the 2-digit occupation “Transportation and Material Moving”, although being very much exposed to robotic labour-saving technologies, has experienced a positive employment growth (cf. Figure 1). Whenever occupations record a positive growth, notwithstanding their high-exposure rate, this signals that employment dynamics are driven by other sources, primarily demand, which might clearly counterbalance the potential labour-saving traits of advancement in robotics. Notably, aggregation at the industry level highlights a further deindustrialisation of the US economy, since manufacturing is by far the most exposed sector. However, social care and assistance services, as well as education, turn out to be quite high in the exposure ranking. Therefore, not only low-paid manufacturing and logistics workers are exposed but also low-paid service workers. On the contrary, managerial occupations, although falling in terms of employment shares, present the lowest degree of similarity.

A secondary result of our study is that high similarity is quite a rare event: the CPC-task cosine similarity matrix is sparse and high values of similarity are more the exception than the norm. This finding corroborates our procedure, providing results which are not inflated but rather conservatively underestimated, given our very cautious identification of robotic labour-saving patents (cf. Section 4.1). As a consequence, when considering the cumulative fraction of potentially replaceable occupations, the top decile of the similarity distribution involves 8.6% or approximately 12.6 million employees.

The strengths of our approach are, first, the construction of a direct measure of proximity by means of an objective procedure, not resorting to subjective and mutable expert judgements, or alternatively to crowdsourcing, and second, its generality, the measure being constructed on the *entire* set of CPC codes, and only in a second step using a weighting procedure to account for labour-saving technologies. This means that we obtain a similarity measure for the entire technological (CPC) and occupational (O*NET tasks) spectra. Finally, the non linear-nature of the quantile threshold and the sparsity of the matrix add support both in terms of reliability and in terms of labour market prospects.

An initial limitation of the present study is that it takes into account only robotic labour-saving innovations, while labour-saving innovations encompass both other applications of AI technologies such as technological change embodied in machineries, and tools distinct from robots. A second limitation is that we are not able to track adopters of these technologies and we do not know the exact number of workers, in terms of intensive rather than extensive margins, each machine embodying a labour-saving technology is able to replace. This potentially means that if adopters are widespread and the number of their labour units is high, the occupational losses might be much higher than predicted in this work. A third limitation is the lack of a time-varying dimension of the measure, although in Montobbio et al. (2022) we have shown that labour-saving patents are quite persistent in our interval of observation, in contrast with the number of robotic patents which is increasing over time, hinting at persistently stable dynamics in LS heuristics. However, new avenues of research would entail the specification of a time-varying measure allowing for the capture of the evolution of underlying technologies. In terms of occupations/tasks, a time-varying dataset would entail the merging of different O*NET waves, spanning at least two decades and studying within-occupational variability in terms of intensity of tasks performed.

Potential extensions of our study involve, first, the LS identification of other technologies beyond the strictly robotic, such as AI or standard ICT. Second, our measure can be adopted to labour markets other than the US. Third, an application of our indicator at the firm-level constitutes a quite promising avenue of research, aiming at pinpointing the establishments and plants more exposed to labour-saving efforts.

ACKNOWLEDGEMENTS

The authors wish to thank two anonymous referees, participants at the 2021 INFER Annual Conference, the 2021 SIE Annual Conference, the 2021 EAEPE Annual Conference, the 2021 International Schumpeter Society Conference, the 2021 ENEF Conference, the 2022 JRC – European Commission Seminar Series, the 2022 IRPET Seminar Series, and the 2021 UNU-MERIT Seminar Series for helpful comments and insightful suggestions at various stages of this work. Fabio Montobbio, Jacopo Staccioli, and Marco Vivarelli acknowledge support by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca, PRIN-2017 project 201799ZJSN: “Technological change, industry evolution and employment dynamics”. Maria Enrica Virgillito acknowledges support from European Union's Horizon 2020 research and innovation programme under grant agreement No. 822781 “GROWINPRO – Growth Welfare Innovation Productivity”.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Maria Enrica Virgillito https://orcid.org/0000-0001-5958-9647

REFERENCES

Acemoglu, D., & Autor, D. H. (2011). Skills, tasks and technologies: Implications for employment and earnings. In D. Card & O. Ashenfelter (Eds.), *Handbook of labor economics* (Chap. 12, Vol. 4B, pp. 1043–1171). Elsevier.

Acemoglu, D., Autor, D. H., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), s293–s340. https://doi.org/10.1086/718327

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020). Competing with robots: Firm-level evidence from France. *AEA Papers and Proceedings*, *110*, 383–388. https://www.doi.org/10.1257/pandp.20201003

Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, *108*(6), 1488–1542. https://doi.org/10.1257/aer.20160696

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, *33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Acemoglu, D., & Restrepo, P. (2020a). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, *128*(6), 2188–2244. https://doi.org/10.1086/705716

Acemoglu, D., & Restrepo, P. (2020b). The wrong kind of AI? Artificial intelligence and the future of labour demand. *Cambridge Journal of Regions, Economy and Society*, *13*(1), 25–35. https://doi.org/10.1093/cjres/rsz022

Aggarwal, C. C. (2018). *Machine learning for text*. Springer.

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The risk of automation for jobs in OECD countries: A comparative analysis*. OECD Social, Employment and Migration Working Papers No. 189. OECD Publishing.

Autor, D. H., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, *103*(5), 1553–1597. https://doi.org/10.1257/aer.103.5.1553

Autor, D., Salomons, A., & Seegmiller, B. (2022). *New Frontiers: The origins and content of New Work, 1940–2018*. NBER Working Paper 30389. https://www.nber.org/papers/w30389

Brynjolfsson, E., & McAfee, A. (2012). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Digital Frontier Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, *108*, 43–47. https://doi.org/10.1257/pandp.20181019

Calvino, F., & Virgillito, M. E. (2018). The innovation employment nexus: A critical survey of theory and empirics. *Journal of Economic Surveys*, *32*(1), 83–117. https://doi.org/10.1111/joes.12190

Cetrulo, A., & Nuvolari, A. (2019). Industry 4.0: Revolution or hype? Reassessing recent technological trends and their impact on labour. *Journal of Industrial and Business Economics*, *46*(3), 391–402. https://doi.org/10.1007/s40812-019-00132-y

Chiacchio, F., Petropoulos, G., & Pichler, D. (2018). *The impact of industrial robots on EU employment and wages: A local labour market approach*. Bruegel Working Paper 2018/02.

Cirillo, V., Rinaldini, M., Staccioli, J., & Enrica Virgillito, M. (2021). Technology vs. workers: The case of Italy's industry 4.0 factories. *Structural Change and Economic Dynamics*, *56*, 166–183. https://doi.org/10.1016/j.strueco.2020.09.007

Dauth, W., Findeisen, S., Südekum, J., & Woessner, N. (2021). The adjustment of labor markets to robots. *Journal of the European Economic Association*, *19*(6), 3104–3153. https://doi.org/10.1093/jeea/jvab012

Dingel, J. I., & Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, *189*, 104235. https://doi.org/10.1016/j.jpubeco.2020.104235

Domini, G., Grazzi, M., Moschella, D., & Treibich, T. (2021). Threats and opportunities in the digital era: Automation spikes and employment dynamics. *Research Policy*, *50*(7), 104137. https://doi.org/10.1016/j.respol.2020.104137

Dosi, G., & Mohnen, P. (2019). Innovation and employment: An introduction. *Industrial and Corporate Change*, *28*(1), 45–49. https://doi.org/10.1093/icc/dty064

Felten, E., Raj, M., & Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. *AEA Papers and Proceedings*, *108*, 54–57. https://doi.org/10.1257/pandp.20181021

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, *42*(12), 2195–2217. https://doi.org/10.1002/smj.3286

Fillmore, I., & Hall, J. D. (2021). Technological change and obsolete skills: Evidence from men's professional tennis. *Labour Economics*, *73*, 102051. https://doi.org/10.1016/j.labeco.2021.102051

Ford, M. (2015). *The rise of the robots. Technology and the threat of mass unemployment*. Basic Books.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Handel, M. J. (2016). The O*NET content model: Strengths and limitations. *Journal for Labour Market Research*, *49*(2), 157–176. https://doi.org/10.1007/s12651-016-0199-8

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, *3*(3), 303–320. https://doi.org/10.1257/aeri.20190499

Koch, M., Manuylov, I., & Smolka, M. (2021). Robots and firms. *The Economic Journal*, *131*(638), 2553–2584. https://doi.org/10.1093/ej/ueab009

Kogan, L., Papanikolaou, D., Schmidt, L. D. W., & Seegmiller, B. (2021). *Technology-skill complementarity and labor displacement: Evidence from linking two centuries of patents with occupations*. Mimeo. https://www.bryanseegmiller.com/files/Draft_v2021-1017.pdf

Krzywdzinski, M. (2021). Automation, digitalization, and changes in occupational structures in the automobile industry in Germany, the United States, and Japan: A brief history from the early 1990s until 2018. *Industrial and Corporate Change*, *30*(3), 499–535. https://doi.org/10.1093/icc/dtab019
Meindl, B., Frank, M. R., & Mendonça, J. (2021). *Exposure of occupations to technologies of the fourth industrial revolution*. arXiv preprint. arXiv:2110.13317.
Mondolo, J. (2022). The composite link between technological change and employment: A survey of the literature. *Journal of Economic Surveys*, *36*(4), 1027–1068. https://doi.org/10.1111/joes.12469
Montobbio, F., Staccioli, J., Virgillito, M. E., & Vivarelli, M. (2022). Robots and the origin of their labour-saving impact. *Technological Forecasting and Social Change*, *174*, 121122. https://doi.org/10.1016/j.techfore.2021.121122
Nedelkoska, L., & Quintini, G. (2018). *Automation, skills use and training*. OECD Social, Employment and Migration Working Papers, No. 202. OECD Publishing.
Santarelli, E., Staccioli, J., & Vivarelli, M. (2022). Robots, AI, and related technologies: A mapping of the new knowledge base. *Journal of Technology Transfer*, *48*, 779–813. https://doi.org/10.1007/s10961-021-09914-w
Staccioli, J., & Virgillito, M. E. (2021). Back to the past: The historical roots of labor-saving automation. *Eurasian Business Review*, *11*(1), 27–57. https://doi.org/10.1007/s40821-020-00179-1
Webb, M. (2020). *The impact of artificial intelligence on the labor market*. https://ssrn.com/abstract=3482150

How to cite this article: Montobbio, F., Staccioli, J., Virgillito, M. E., & Vivarelli, M. (2023). Labour-saving automation: A direct measure of occupational exposure. *The World Economy*, *00*, 1–30. https://doi.org/10.1111/twec.13522

APPENDIX 1

In this Appendix, we briefly summarise the methodological steps adopted in Montobbio et al. (2022).

Step 1 – Identification of robotic patents

The analysis starts with the entire set of 3,557,435 patent applications published by the USPTO between 1st January 2009 and 31st December 2018. Robotic patents are pinpointed therein according to two distinct criteria, one based on the patent classification codes specified within applications, the other based on a textual keyword search. A patent is deemed 'robotic' if it obeys at least one of the criteria. In particular, a robotic patent according to the first criterion (dubbed 'CPC') must be assigned at least one of a set of 174 full-digit CPC codes which reflect former US Patent Classification (USPC) class 901 ("Robots") by patent examiners. According to the second criterion (dubbed 'K10'), a robotic patent must contain the word 'robot' in its full text at least 10 times, including derivational and inflectional affixes. The first criterion identifies 10,929 robotic patents, while the second criterion identifies another 18,860 (after discarding robotic patents according to the first criterion). The two criteria single out a total of 29,789 robotic patents, i.e., approximately 0.84% of the original (universe) population.

Step 2 – Identification of labour-saving patents

LS patents constitute a subset of robotic patents, identified by a multiple *word* co-occurrence query at the *sentence* level. In particular, a patent is deemed LS (after an additional manual validation step) if its full text contains at least one sentence in which the verbal predicate, direct object and object attribute belong to the following lists:

$$\underbrace{\begin{bmatrix}\text{'reduc'}\\\text{'replac'}\\\text{'elimin'}\\\text{'save'}\\\text{'lower'}\\\text{'substitut'}\\\text{'autom'}\end{bmatrix}}_{\text{verbal predicate}} \times \underbrace{\begin{bmatrix}\text{'labor'}\\\text{'worker'}\\\text{'human'}\\\text{'employe'}\\\text{'manpow'}\\\text{'job'}\end{bmatrix}}_{\text{direct object}} \times \underbrace{\begin{bmatrix}\text{'cost'}\\\text{'expenditure'}\\\text{'expens'}\\\text{'hour'}\\\text{'intens'}\\\text{'task'}\\\text{'time'}\\\text{'skill'}\end{bmatrix}}_{\text{object attribute}}$$

In total, 1276 LS patents are found (approximately 4.3% of all robotic patents), of which 461 (≈36.1%) belong to the CPC group and 815 (≈63.9%) belong to the K10 group.

APPENDIX 2

Each piece of text, either a CPC definition or an O*NET task description, first undergoes a pre-processing step in which words are *stemmed* to their morphological root and so-called stop-words, i.e., tokens that are overly common in English (such as 'a', 'the', 'if'...) and do not convey any useful information to our analysis, are removed. Each text is then transformed into a vector of frequencies of the underlying words. The number of vector components reflects the common dictionary of terms across the two whole corpora. In other words, all vectors belong to the same vector space, whose dimension equals the number of distinct words in the common dictionary. The similarity of each CPC-task pair is then quantified as the cosine of the angle between the two underlying vectors.

As opposed to simply counting the occurrences of each word in each body of text, we adopt the customary *tf-idf* (term frequency–inverse document frequency) term-weighting scheme for computing the relevant frequencies, according to the following Definition.

Definition 1. Let D be a collection of documents d, each composed of an ensemble of terms from a dictionary T. The tf-idf measure of term t appearing in document d is defined as follows:

$$\text{tf-idf}(t, d, D) := \text{tf}(t, d) \cdot \text{idf}(t, D), \quad \forall d \in D, \forall t \in T$$

$$\text{tf}(t, d) := 1_d(t) = \begin{cases} 1 & \text{if } t \in d \\ 0 & \text{otherwise} \end{cases}, \quad \forall d \in D, \forall t \in T$$

$$\text{idf}(t, D) := \log\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right), \quad \forall t \in T$$

The associated $|D| \times |T|$ document term matrix $\mathcal{D}^D$ is an array of tf-idf measures for all documents d in the generic collection D and for all terms t in the relevant dictionary T. In other words,

$$\mathcal{D}^{D}_{d,t} = \text{tf-idf}(t, d, D), \quad \forall d \in D, \forall t \in T$$

The tf-idf statistic reflects how important a specific term is to a certain document, compared to other documents in the collection. The tf-idf value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus which mention that word. This helps to adjust for the fact that some words appear more frequently in general.

Extending the reasoning to the corpus level, we construct two *document-term* matrices, $\mathcal{D}^{CPC}$ and $\mathcal{D}^{TASK}$, whose rows contain the aforementioned tf-idf frequency vectors, for each CPC code definition and each task description, respectively. Both matrices are based on the dictionary of terms from CPC definitions, namely the smaller between the two collections, which consists of 2309 terms. Therefore $\mathcal{D}^{CPC}$ has dimension 671×2309 and $\mathcal{D}^{TASK}$ has dimension $19{,}231 \times 2309$. Finally, we construct the cosine similarity matrix S containing the cosine similarity score between all pairs of row vectors from the document-term matrices $\mathcal{D}^{CPC}$ and $\mathcal{D}^{TASK}$ according to the following Definition.

Definition 2. Given two vectors $X, Y \in \mathbb{R}^{|T|}$, their cosine similarity is defined as the cosine of the angle between them, which is also equal to the inner product of the same vectors normalised to unit length, as follows:

$$cos(X, Y) := \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{\sum_{t=1}^{|T|} x_t y_t}{\sqrt{\sum_{t=1}^{|T|} x_t^2} \sqrt{\sum_{t=1}^{|T|} y_t^2}}$$

where x_t and y_t denote the components of vectors X and Y, respectively, and $\| \cdot \|$ denotes the Euclidean norm.

Since row vectors of document-term matrices are non-negative valued, their cosine similarity is bounded by the unit interval, i.e. $cos(X, Y) \in [0, 1]$. Moreover, when term frequency is measured by tf-idf, the normalisation denominator in the equation above is redundant and $cos(X, Y) \equiv X \cdot Y$. Therefore, given document-term matrices $\mathcal{D}^{CPC}$ and $\mathcal{D}^{TASK}$, and extending the cosine similarity computation to the matrix level, the following cosine similarity matrix is obtained:

$$S = cos\left(\mathcal{D}^{CPC} \cdot \mathcal{D}^{TASK}\right) \equiv \mathcal{D}^{CPC}\left(\mathcal{D}^{TASK}\right)'.$$

APPENDIX 3

In order to test the robustness of our procedure, we replicate the text similarity exercise described in Section 4 using LS patent full texts, rather than CPC code definitions. In principle, one might expect the underlying content of the patent to be more informative than CPC definitions when it comes to actual labour-saving efforts and any human functions substituted.

Tables A1 and A2 present the results of the top-task and top-occupation matching by similarity scores. Interestingly, the identified tasks and occupations are completely different from the original exercise. Emerging tasks are "Build or assemble robotic devices or systems"; "Set up and operate computer-controlled machines or robots to perform one or more machine functions on metal or plastic workpieces"; "Build, configure, or test robots or robotic applications";

TABLE A1 Top 20 tasks by (rescaled) cosine similarity (rounded to the 2nd decimal digit) based on LS patents full texts.

Rank	Code	Description	Cosine similarity
1	16596	Build or assemble robotic devices or systems	1.0
2	11944	Set up and operate computer-controlled machines or robots to perform one or more machine functions on metal or plastic workpieces	0.98
3	21057	Build, configure, or test robots or robotic applications	0.97
4	16523	Conduct research on robotic technology to create new robotic systems or system capabilities	0.93
5	16511	Provide technical support for robotic systems	0.91
6	16587	Assist engineers in the design, configuration, or application of robotic systems	0.86
7	16525	Conduct research into the feasibility, design, operation, or performance of robotic mechanisms, components, or systems, such as planetary rovers, multiple mobile robots, reconfigurable robots, or man-machine interactions	0.84
8	16593	Install, program, or repair programmable controllers, robot controllers, end-of-arm tools, or conveyors	0.81
9	16584	Modify computer-controlled robot movements	0.80
10	16579	Maintain service records of robotic equipment or automated production systems	0.80
11	20262	Plan special events, parties, or meetings, which may include booking musicians or celebrities	0.80
12	14861	Inquire about pesticides or chemicals to which animals may have been exposed	0.79
13	16075	Implement controls to provide security for operating systems, software, and data	0.79
14	23748	Prepare and submit reports that may include the number of passengers or trips, hours worked, mileage, fuel consumption, or fares received	0.79
15	1277	Perform systems analysis and programming tasks to maintain and control the use of computer systems software as a systems programmer	0.79
16	2463	Develop networks of attorneys, mortgage lenders, and contractors to whom clients may be referred	0.78
17	246	Arrange for medical, psychiatric, and other tests that may disclose causes of difficulties and indicate remedial measures	0.77
18	11338	Transport mail from one work station to another	0.77
19	18280	Install, calibrate, or maintain sensors, mechanical controls, GPS-based vehicle guidance systems, or computer settings	0.77
20	16434	Calibrate vehicle systems, including control algorithms or other software systems	0.77

TABLE A2 Top 20 occupations by (rescaled) cosine similarity (rounded to the 2nd decimal digit) based on LS patents full texts.

Rank	Code	Title	Cosine similarity
1	17-2199.08	Robotics Engineers	1.0
2	17-3024.01	Robotics Technicians	0.96
3	47-2231.00	Solar Photovoltaic Installers	0.49
4	17-2072.01	Radio Frequency Identification Device Specialists	0.46
5	15-1299.08	Computer Systems Engineers/Architects	0.45
6	15-1299.02	Geographic Information Systems Technologists and Technicians	0.42
7	51-9161.00	Computer Numerically Controlled Tool Operators	0.41
8	17-2199.11	Solar Energy Systems Engineers	0.40
9	49-2091.00	Avionics Technicians	0.39
10	15-1243.01	Data Warehousing Specialists	0.38
11	17-1022.01	Geodetic Surveyors	0.38
12	15-1244.00	Network and Computer Systems Administrators	0.38
13	17-2061.00	Computer Hardware Engineers	0.37
14	15-1299.03	Document Management Specialists	0.37
15	15-1211.00	Computer Systems Analysts	0.36
16	51-4034.00	Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.36
17	17-2041.00	Chemical Engineers	0.36
18	49-9044.00	Millwrights	0.36
19	15-2051.02	Clinical Data Managers	0.36
20	17-3021.00	Aerospace Engineering and Operations Technologists and Technicians	0.35

"Conduct research on robotic technology to create new robotic systems or system capabilities"; "Provide technical support for robotic systems". These tasks are clearly labour-complementing, i.e., required to develop and manufacture the new robotic artefacts. Occupations more exposed to labour-complementarity are indeed "Robotics Engineers"; "Robotics Technicians"; "Computer Systems Engineers/Architects"; "Data Warehousing Specialists"; "Network and Computer Systems Administrators".

Notably, the similarity measure by occupations presents a drop after the first two most-exposed occupations onwards. In addition, the elicited occupations reveal who those workers programming and creating LS technologies are: they belong to the upper echelon of the occupational categories, are well paid, and have grown in number during the last two decades. Indeed, top-complementary occupations tend to belong to "Computer and Mathematical Occupations" (15) and "Architecture and Engineering Occupations" (17).